

BRAZILIAN JOURNAL OF BIOMETRICS

ISSN:2764-5290

ARTICLE

Probabilistic Modeling of the Coffee Market in Brazil

 Otávio Magno Teixeira ¹,  Jerônimo Alves Santos ²,  Gilberto Rodrigues Liska ²

¹ Undergraduate student in Agricultural Engineering, Federal University of Sao Carlos, Araras - SP, Brasil.

² Department of Agroindustrial Technology and Rural Socioeconomics, Federal University of Sao Carlos, Araras - SP, Brasil

(Received: August 14, 2023; Revised: October 24, 2023; Accepted: November 27, 2023; Published: April 15, 2024)

Abstract

The coffee is one of the most valuable commodities in the world and in this scenario, Brazil presents itself as the largest producer and exporter in the world. However, the high fluctuations in prices promote insecurity in the agents of the sector. In this sense, the objective of this work is to propose the best probabilistic model for the monthly analysis of prices and to calculate probabilities of occurrences of average prices of coffee bags according to levels of practical interest. To this end, we use historical data provided by COOXUPÉ corresponding to the period from January 1981 to December 2022, arranged into monthly subseries. For supporting the results, goodness of fit test were performed. The results indicated that the Gamma and log-Normal distributions fit the coffee bag price data in all months. The log-Normal distribution outperformed in all months. The Gamma and log-Normal distributions fitted the monthly data of average coffee bags prices. The Log-Normal distribution is more suitable for the probabilistic study of the variable in all months. January, February, and March are the months with the highest probability of higher average values and are therefore the most recommended for coffee trading.

Keywords: Economic series, R programming language, Commodity, Probability, Coffee prices.

1. Introduction

Coffee is one of the most valuable commodities in the world, and approximately 70% of the global production is traded internationally (ICO 2020; Paiva *et al.* 2018). According to the International Coffee Organization (2021), this is a growing market, with global production volume having increased by over 60% since the 1990s and now holding an estimated value of over USD 200 billion annually. The coffee's notable participation in the global economy reflects the influence of the commodity on the economies of various emerging producing countries. In countries with minimal development, exports can contribute to foreign exchange, highlighting the impact on economic and political structure, as well as the risks of collapse if a permanent price decline occurs (Sá Barreto & Zugaib, 2016).

Brazil is the world's largest producer and exporter of coffee, and the coffee activity has a

significant role in the country's economy and national agricultural production (Carrasco-Gutierrez, Almeida, 2013; CONAB, 2022; Mehta, Chavas, 2008). According to the MAPA (2022), coffee moved USD 6.4 billion in 2021, accounting for 8.28% of the country's exports. However, the high market volatility causes insecurity for coffee growers, as fluctuations impact the price of the final product and the economic returns of the activity (Arêdes *et al.*, 2009; Sá Barreto & Zugaib, 2016).

Being a commodity, coffee is one of the agricultural products that is primarily traded in its raw or semi-processed form. Globally, its price is standardized with quotations and negotiations conducted through commodity exchanges (Albieri & Terra, 2022). According to Sá Barreto & Zugaib (2016), price fluctuations are primarily linked to supply and demand movements, as experienced throughout history. An excess of supply leads to price decreases, while scarcity results in price increases. Nevertheless, there are other non-systemic and market-related factors that influence this determination and shape production cycles. The authors mention among market agents the actions of investment funds, exchange rates, market speculation, interest rates, economic crises, and the existence of possible others. For non-systemic agents, climate changes were highlighted, a factor closely related to production.

Furthermore, the coffee market is characterized by oligopsony, where the supply comes from many producers, but the demand is concentrated among a small number of buyers (Ghoshray & Mohan, 2021; Sá Barreto & Zugaib, 2016). This market concentration enhances the influence of coffee buyers from importing countries on prices, while diminishing such capability in producing countries (Ghoshray & Mohan, 2021). Another important factor for price dynamics is that the activity was impacted by the Covid-19 pandemic; the uncertainties of the period affected both supply and demand in the sector while causing disruptions throughout the production chain (ICO, 2020). Currently, coffee production is being influenced by the military conflict involving Russia and Ukraine. In 2020, these countries accounted for 3.85% of global coffee consumption. Russia supplies 20% of the ammonia transported by sea in the global market, and the tension caused an increase in oil prices (ICO, 2022).

Understanding the dynamics of the coffee market enables better risk management, allowing producers to make informed decisions in cultivation and position their product in the market with optimal pricing and profitability (Carrasco-Gutierrez & Almeida, 2013). This understanding can also support the formulation of political and economic strategies by governments of exporting and importing countries, cooperatives, futures market participants, roasters, and others (Carrasco-Gutierrez & Almeida, 2013; Ghoshray & Mohan, 2021; Sá Barreto & Zugaib, 2016). The application of statistical and econometric models in studying the dynamics of the sector is widely spread. The employed model will vary depending on the intended objective. For example, if the purpose is to understand the causal structure of the market, it is possible to estimate and test structural equations. However, if the goal is to test hypotheses among variable correlations, forecasting, or even formulating policies, it is possible to perform partial or complete sector modeling (Carrasco-Gutierrez & Almeida, 2013).

Given the current market scenario, its significance, and the disruptions caused by the Covid-19 pandemic and the conflicts between Ukraine and Russia, understanding the dynamics of coffee prices is of substantial importance. In the present study, this understanding will be achieved through probabilistic models, enabling the development of strategies by all market agents involved, within the context of price volatility and its related consequences. To achieve the proposed objective, the following steps were taken: i) Which probabilistic model is the most appropriate for modeling the average sales prices for bags of coffee in Brazil?; and ii) Which probability(s) of occurrence of a price of bags coffee be greater than a value of practical interest? What is the expected price(s) such that the probability of this being exceeded occurs with low probability?

2. Materials and Methods

2.1 Data description

The series of historical data obtained by COOXUPÉ for 41 years was analyzed, covering the period from January 1981 to December 2022, consisting of records of coffee bag prices at the national level was followed. The time series consist of monthly information about the average price of coffee bags traded at COOXUPÉ in dollars. In keeping with the proposed objectives, the data series was divided into monthly subseries, totaling twelve subseries. Each subseries was split into two parts, one called the training set and dedicated to fitting the probability distributions and the other series, called the test set, dedicated to measuring the goodness of fit. The training dataset covered 31 years (1981 to 2012), around 75% of the data, and the test dataset the last 10 years (2013 to 2022), in accordance with what is recommended by Menezes *et al.* (2017).

It is worth noting that COOXUPÉ's arabica coffee production is exported to 49 countries across five continents. In the coffee export ranking, the cooperative has been the top exporter among operating companies in Brazil for seven consecutive years. Exports account for more than 60% of COOXUPÉ's activities (COOXUPÉ, 2022). The decision to use dollar prices was based on the intention of mitigating the effects of changes in monetary policies that occurred during the analysis period. It is important to highlight that the data obtained represent the average price of the values passed on over the years to cooperators, therefore, it reflects the average amount delivered per bag of coffee to each cooperator, regardless of the variety of coffee produced. In this context, it is noted that these values only include the incomes received from coffee production by cooperators, which are rural producers.

To follow the data description, descriptive statistics were calculated for the data, encompassing measures such as maximum value, minimum value, mean, median, standard deviation, coefficient of variation, and skewness coefficient.

2.2 Probabilistic modeling of the price of coffee bags

Let the month m (m =January, ..., December), and the respective training dataset, the probability distributions used in this study for each m will be presented below. Let X the random variable associated with the price of a coffee bags (in dollars), the Gamma distribution has a probability density function given by:

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (1)$$

defined in $X \geq 0$, β is the scale parameter ($\beta > 0$), α is the shape parameter ($\alpha > 0$), and the gamma function is defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. The probability density function of the Log-Normal distribution is given by:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}} \quad (2)$$

where $X > 0$, \ln represents the natural logarithm, μ ($-\infty < \mu < \infty$) is the mean of the natural logarithms of variable X , and σ ($\sigma > 0$) is the standard deviation of the natural logarithms of random variable X . The Normal distribution has a probability density function given by:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3)$$

where $-\infty < X < \infty$, μ ($-\infty < \mu < \infty$) is the mean and σ^2 is the variance of the random variable X (Devore, 2018).

The probabilistic models in equations 1 and 2 have asymmetric shape, and the one in equation 3 has symmetric shape. The parameters of these functions are initially unknown. To determine them, some estimation method is required. For this purpose, the maximum likelihood method was employed, which involves determining the values of the parameters α , β , μ and σ that maximize the probability of the sample having occurred, under the condition of independence. This fact was verified by the Ljung Box and Mann-Kendal tests and more details are in section 2.3. Further details about the maximum likelihood method can be found in Liska *et al.* (2020).

With each of the above models, we can proceed with the calculation of probabilities and/or quantiles of practical interest. The probability of a coffee sack's average price occurring up to a hypothetical value y is calculated using the cumulative distribution function ($F(y)$), which means that

$$F(y) = P[X \leq y] = \int_{-\infty}^y f(x; \theta) dx, \quad (4)$$

in other words, the probability of an average price occurring up to a quantity y is given by the integral solution of any of the probability density functions, $f(x; \theta)$ which θ is a parameter or vector of parameters, presented earlier. Therefore, the probability of an average coffee sack price exceeding a value y is given by:

$$P[X > y] = 1 - F(y). \quad (5)$$

The calculation of quantiles of practical interest is obtained by taking the inverse of the cumulative distribution function given the probability. Further details can be found in Casella and Berger (2001).

2.3 Hypothesis testing and goodness of fit

After the proposal of a probabilistic model and the estimation of its parameters, it is necessary to assess whether the proposed model fits the data. This is done through a goodness of fit test and it was done with the training set, according explained on the 2.1 section. The main idea of an adherence test is to compare the empirical probabilities with the theoretical probabilities predicted by the distribution function under test, checking whether the sample values can be found from a population with that distribution (Hartmann *et al.*, 2011). Among them, there is the Kolmogorov-Smirnov (KS) test, which is based on the analysis of proximity or adjustment between the sampling distribution function and the population distribution function. The test statistic is given by,

$$D = \max |F(x_{(i)}) - \hat{F}(x_{(i)})|, \quad i = 1, 2, \dots, n \quad (6)$$

where $F(x_{(i)})$ being the theoretic cumulative distribution function of the densities (1), (2) e (3) with the respectively parameters estimates, $\hat{F}(x_{(i)})$ is the empirical cumulative distribution function and n the length do training set. In this test, the null hypothesis claims whether the distribution function from which the sample comes from follows the distribution function that is assumed to be known, and the alternative hypothesis the opposite. Then, value (6) must be compared with the critical value (using tables), for the level of significance of the test. Depending on the result, the null hypothesis is rejected or not. Strictly speaking, the KS test assumes that the theoretical

distribution function refers to a population and, consequently, its parameters. However, its use with parameter estimates has been widely used and, as reported by Bautista *et al.* (2004), the type 2 error rates can be inflated. To get around this fact, the Chi-Square also was applied and it compares the observed and expected cumulative frequencies using the test statistic:

$$Q = \sum_{i=1}^k \frac{(\hat{F}(x_{(i)}) - F(x_{(i)}))^2}{F(x_{(i)})} \quad (7)$$

where k is the number of bins, calculated according Sturges rule, where $k = 1 + 3.222 \ln n$, $\hat{F}(x_{(i)})$ is the observed cumulative frequency of the i -th bin, and $F(x_{(i)})$ is the expected frequency of the i -th bin according to the distribution being tested. The hypotheses of the chi-squared test are the same as those of the KS test. The Q statistic is evaluated using the chi-squared distribution with $k-1$ degrees of freedom, and if $Q > \chi_{\alpha}^2(k-1)$, where α is the significance level, the null hypothesis of the test is rejected. In both tests, the null hypothesis is also rejected if the p -value is lower than the adopted significance level. The Ljung Box (LB) independence test, whose statistics are compared with the α -th quantile of the chi-squared distribution with one degree of freedom. If the LB test statistic is higher than the aforementioned quantile, the null hypothesis of independence of the training set is rejected. Alternatively, if the p -value is smaller than the adopted significance level, the null hypothesis is rejected. The Mann-Kendall test was used to determine if the series has a statistically significant time trend (Sá *et al.*, 2018). When very small values of p -value are found, it indicates evidence in favor of the alternative hypothesis, that is, there is some tendency to modify the behavior of the analyzed series. For all tests, a significance level α of 1% were used.

In order to evaluate the predictive capacity of the fitted models based on the test set, as described in section 2.1, the expected return prices were calculated. With the expressions in (4) and (5), the expected price can be obtained given a return time T . The return time T represents the inverse of the probability with which a given event E has occurred. Given the occurrence of an event E , the return time T is the average time required (in days, months or years) for this event to recur, at any given time. By definition, it follows that the return time T associated with event E is given by

$$T = \frac{1}{P[E]}. \quad (8)$$

In the present work, event E is the price of a bag of coffee (in dollars) exceeding a certain value and the probability of exceeding E is obtained by $1 - F(y)$. So

$$T = \frac{1}{P[E]} = \frac{1}{1-F(y)}. \quad (9)$$

For simplicity, consider $P[E] = p$. The return level (z_t) is obtained by the solution of the following equation $\int_{-\infty}^{z_t} f(x; \theta) dx = 1 - p$, where z_t is related to the return time t and $p = \frac{1}{t}$. By equation (4), it is noticed that $F(z_t) = \int_{-\infty}^{z_t} f(x; \theta) dx = 1 - p$, so, as $F(y)$ has an inverse function (F^{-1}), the return level can be calculated, i.e. $z_t = F^{-1}(1 - p)$. In this way, with the estimates of the parameters of the probabilistic models, a prediction of the return price for the return time t is obtained, given by

$$\hat{z}_t = F^{-1}(1 - p) \quad (10)$$

Given the above, the goodness of fit of the models was assessed in comparisons of the empirical return level (z_t) with the expected return level (\hat{z}_t) and considering the return times of 2, 3, 4, 5, 6, 7, 8, 9 and 10 years. The mean absolute percentage error (MAPE) was used, given by

$$MAPE = \frac{1}{n} \sum_{i=1}^{n_t} \left| \frac{z_{t_i} - \hat{z}_{t_i}}{z_{t_i}} \right| \times 100 \quad (11)$$

where, n_t is the number of predictions made, in this case $n_t=10$. The model that presents the lowest MAPE is evidence in favor of it.

Concluding the methodology, the statistical computational system R (R Core Team, 2021) was used for conducting the analyses with help of the packages MASS (Venables & Ripley, 2002), hydroGOF (Zambrano-Bigiarini, 2020) and trend (Pohlert, 2023).

3. Results and Discussion

The results shown in Table 1 present some descriptive statistics related to the data rearranged in monthly series, as mentioned in Section 2.1. It is possible to observe that the highest values recorded for the maximum of average prices per sack occurred in the months of April, May, and January, respectively, while the lowest values recorded for the minimums were in the months of October, June, and July.

The highest standard deviations of the commodity were observed in the months of January, February, and March, in that order. This measure of data dispersion is traditionally used in determining price volatility and considered a risk measure by some market speculators (Capitani & Mattos, 2017).

Among the monthly subseries, the months of February, January, and March, respectively, had the highest means, while the lowest means occurred for the months of October, September, and November. These values may be correlated with the phenological and physiological cycle of the coffee plant, combined with market reactions. According to Sá Barreto and Zugaib (2016), one of the main factors contributing to price instability is production and supply situation.

Table 1. Descriptive measures of the monthly average prices of coffee bags (in USD) traded at COOXUPÉ from January 1981 to December 2021 and p -values of the Mann-Kendal (MK) and Ljung-Box (LB) tests

Month	Minimum	Maximum	Median	Mean	Standard Deviation	Coefficient of Variation	Coefficient of Asymmetry	MK	LB
January	43.22	309.97	112.35	118.36	64.03	0.5410	1.3646	0.0719	0.0701
February	43.08	284.31	104.84	118.80	63.46	0.5341	1.2365	0.0771	0.0234
March	47.59	297.90	104.20	116.68	60.28	0.5167	1.2884	0.0886	0.0255
April	48.52	319.37	98.02	113.18	59.34	0.5243	1.5128	0.0497	0.0237
May	39.34	310.14	98.84	112.06	58.94	0.5260	1.3812	0.0669	0.0275
June	35.21	305.53	97.57	108.07	56.96	0.5270	1.4272	0.0286	0.0179
July	35.63	284.72	86.75	106.57	59.23	0.5558	1.2118	0.0187	0.0168
August	35.78	277.05	84.43	107.03	59.23	0.5534	1.0274	0.0143	0.0128
September	36.06	261.90	81.76	106.27	58.57	0.5512	1.0048	0.0171	0.0188
October	35.04	259.27	89.73	103.72	54.40	0.5244	1.0297	0.0222	0.0225
November	35.74	267.47	96.54	106.34	54.86	0.5159	0.9883	0.0365	0.0210
December	39.03	259.41	102.57	111.52	56.78	0.5091	0.9456	0.0336	0.0194

Camargo and Camargo (2001) outlined the phenology of coffee plants (*Coffea arabica* L.) for Brazilian tropical conditions and noted in their work that during the flowering period, which occurs from September to December of the first year and in months preceding the rise in average prices, the incidence of high temperatures and an intense water deficit at its beginning leads to flower abortion. Subsequently to fertilization, there are the stages of small green berries and expansion. During these stages, a severe drought can hinder fruit growth and result in a low-quality harvest. In the grain formation stage, which occurs from January to March of the second year, corresponding to months with higher average prices, water stress leads to the production of poorly formed fruits and drying degree (poor development) of grains (Camargo & Camargo, 2001). Coffee harvesting takes place during the dry months, typically between March and April to September, a period when average prices tend to decline.

Beyond that, Silveira, Mattos, and Saes (2016) discuss the importance of crop reports and

their impacts on the market and future price volatility. This is a significant factor considering that coffee trade mainly occurs in commodity exchanges and futures markets. Reports released by the International Coffee Organization, especially during flowering periods, have a significant impact on prices. Similar effects, albeit on a smaller scale, are observed in studies published by CONAB (National Supply Company) in September and December.

From Table 2, it can be observed that the Gamma and log-Normal distributions fitted the coffee bags price data for all months according to the Kolmogorov-Smirnov and χ^2 tests at a 1% significance level. Given the aforementioned and considering that both distributions are asymmetric, a similar result was found for the commodity price of fat ox in the state of São Paulo by Lucca Filho *et al.* (2022). In addition to the results of the adherence tests, it is possible to verify in the histograms in Figure 1 that the probability densities analyzed had a satisfactory fit to the data. Therefore, to choose the most appropriate probability distribution in a given month, the combination of results from Table 2 and Figure 1 must be considered.

When analyzing the calculated skewness coefficients for the monthly subseries, a positive skewness is observed in all months. Therefore, asymmetric distributions are the most suitable for monthly modeling. The predilection for a particular distribution in some months arises from a slightly better fit. Both goodness of fit tests have their limitations, and in some cases, the KS and χ^2 tests selected different distributions, making the MAPE a third criterion for choice. In this way, the log-Normal distribution performed better in all months.

Table 2. Parameter estimates for the probabilistic models, results of the Kolmogorov-Smirnov (KS) and Chi-Square (χ^2) tests, and the Mean Absolute Percentage Error (MAPE) values for the monthly subseries of coffee bags average prices in dollars

Month	Distribution ⁽¹⁾	Parameter ⁽²⁾				KS ^(2,3)	χ^2 ^(2,3)	MAPE ⁽⁴⁾
		α	β	μ	σ			
January	Gamma	4.2127	0.0356	-	-	0.8150	0.3937	33.8434
	Normal	-	-	118.3641	63.0236	0.2550	0.0377	38.3103
	log-Normal	-	-	4.6504	0.4883	0.9663	0.5380	32.1427
February	Gamma	4.2626	0.0359	-	-	0.6404	0.2772	32.6343
	Normal	-	-	118.8000	62.4558	0.1714	0.0209	36.7889
	log-Normal	-	-	4.6556	0.4855	0.9281	0.4623	30.9593
March	Gamma	4.5718	0.0392	-	-	0.7979	0.4701	27.1575
	Normal	-	-	116.6775	59.3335	0.2629	0.0356	31.0393
	log-Normal	-	-	4.6461	0.4673	0.9341	0.7102	25.7472
April	Gamma	4.5198	0.0399	-	-	0.9729	0.2280	27.4037
	Normal	-	-	113.1806	58.4065	0.5668	<0,0001	30.3636
	log-Normal	-	-	4.6143	0.4703	0.9798	0.5547	26.5523
May	Gamma	4.3479	0.0388	-	-	0.9397	0.3518	29.8690
	Normal	-	-	112.0556	58.0116	0.7349	0.0008	32.2862
	log-Normal	-	-	4.5996	0.4841	0.9602	0.5599	29.3095
June	Gamma	4.3097	0.0399	-	-	0.9342	0.3105	26.7420
	Normal	-	-	108.0725	56.0614	0.6921	0.0001	28.9414
	log-Normal	-	-	4.5623	0.4883	0.8801	0.4307	26.4454
July	Gamma	3.9256	0.0368	-	-	0.4978	0.2193	23.8603
	Normal	-	-	106.5656	58.3005	0.1804	0.0339	26.1882
	log-Normal	-	-	4.5360	0.5066	0.7919	0.2229	22.9961
August	Gamma	3.7483	0.0350	-	-	0.3958	0.7722	27.6247
	Normal	-	-	107.0291	58.3017	0.1249	0.2679	29.1757
	log-Normal	-	-	4.5338	0.5262	0.6850	0.7479	27.3773
September	Gamma	3.8191	0.0359	-	-	0.3409	0.6431	32.1040
	Normal	-	-	106.2662	57.6493	0.1101	0.2142	33.9754
	log-Normal	-	-	4.5294	0.5187	0.6039	0.7405	31.5723
October	Gamma	4.1949	0.0404	-	-	0.6414	0.9613	32.7523
	Normal	-	-	103.7237	53.5389	0.4443	0.4508	34.4288

	<i>log-Normal</i>	-	-	4.5178	0.4950	0.7824	0.9728	32.2205
November	Gamma	4.2551	0.0400	-	-	0.6081	0.4907	33.5025
	Normal	-	-	106.3378	53.9920	0.3962	0.2247	34.8023
	<i>log-Normal</i>	-	-	4.5445	0.4933	0.7923	0.3726	33.2217
December	Gamma	4.2671	0.0383	-	-	0.9097	0.5815	35.7566
	Normal	-	-	111.5206	55.8843	0.6978	0.0891	36.6612
	<i>log-Normal</i>	-	-	4.5925	0.4969	0.9213	0.7058	36.1512

(1) Probability distributions in italics show the best at least two of the goodness of fit index KS, χ^2 and MAPE; (2) Results obtained through the training set (January 1981 to December 2012); (3) Result refers to the p -value obtained when carrying out the test for training set; (4) Results obtained through the test set (January 2013 to December 2022)

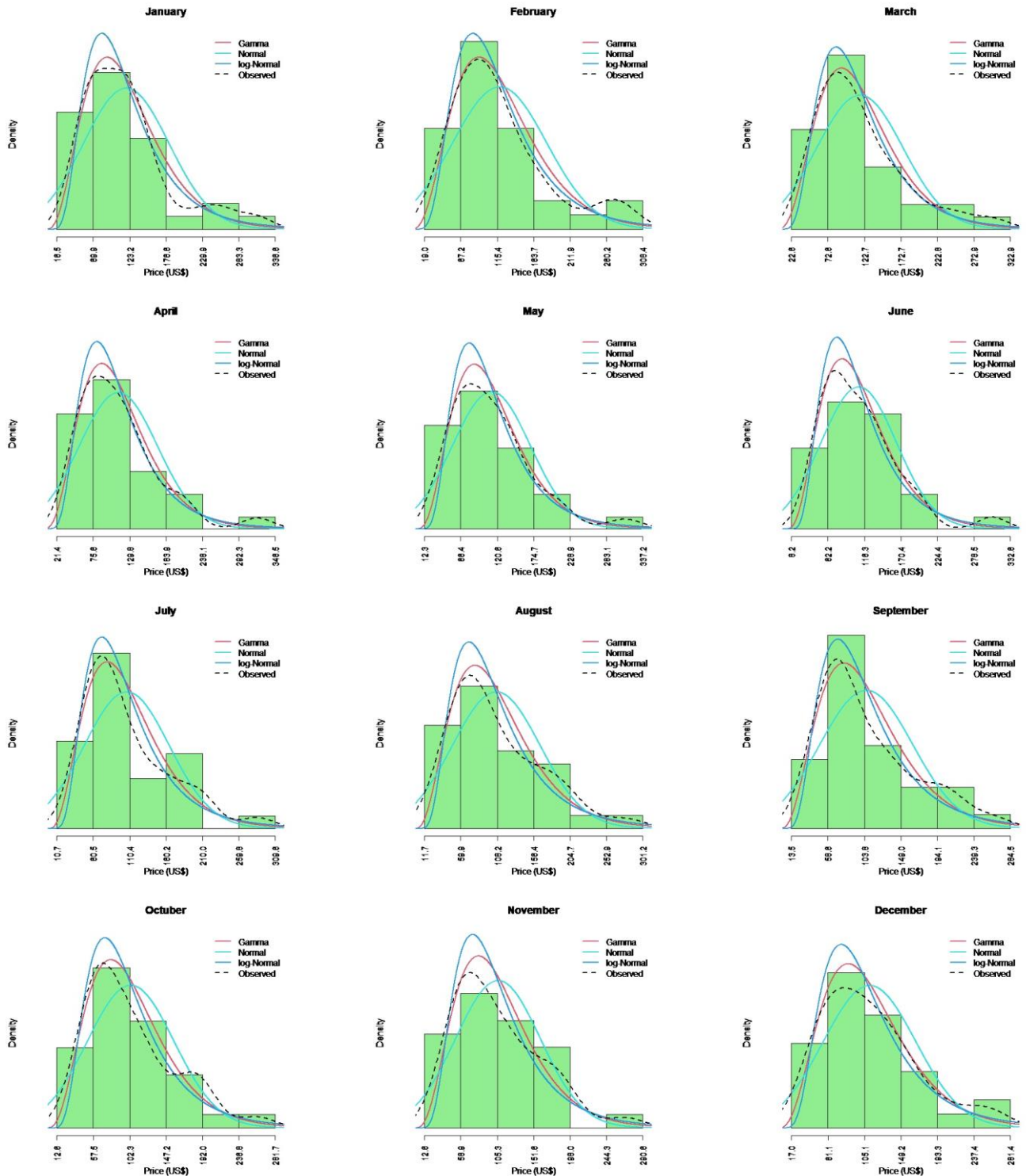


Figure 1. Histograms of the monthly average prices of coffee bags (in USD) from January 1981 to December 2012, along with plots of the probability density functions of the fitted Gamma, Normal, and log-Normal distributions.

In Table 3, it is possible to observe the probabilities of commodity price occurrences and the 1 and 5% quantiles. In the month of January, based on the probabilities calculated using the log-normal distribution identified as the best fit for that month, the probability of prices exceeding 50, 100, 200, and 400 dollars are 93.47%, 53.69%, 9.23%, and 0.30%, respectively. At February, January and March, respectively, showed the highest probabilities for the occurrence of values surpassing 50, 100 and/or 200 dollars. Among all months, January and February are also the months with the highest probability of prices exceeding 400 dollars. On the other hand, among all the months, October has the lowest probability of values exceeding 100, 200 and/or 400 dollars. This trend is like what occurred in the years 2018 and 2019, driven in 2018 by the completion of the harvest and the confirmation of Brazil's record production of 59.4 million bags, coupled with global production exceeding consumption (CONAB, 2018, 2019).

The volatility among prices and monthly variations can be explained by the low short-term elasticity of demand and supply in the coffee market (Sá Barreto & Zugaib, 2016; Silveira *et al.*, 2016), in addition to other factors such as dependency on exchange rates (Carrasco-Gutierrez & Almeida, 2013), macroeconomic, financial, climatic, and meteorological variables (Cuaresma *et al.*, 2018). Furthermore, the monthly series studied were shown to be trendless and independent using the Mann-Kendal and Ljung-Box tests, which makes the parameter estimates obtained using the maximum likelihood method plausible, in terms of bias. If any violation in these tests were found, the incorporation of this aspect in the analysis would be necessary, such as those carried out by Lucca Filho *et al.* (2022) with the incorporation of the trend in probabilistic modeling and Teixeira, Liska and Santos (2023) with the incorporation of temporal modeling of coffee bag prices.

Table 3. Probabilities (in %) of the occurrence of exceeding average prices of coffee bags (in USD) and quantiles of practical interest for all months of the year

Month	Distribution*	Probability				Quantile	
		>\$50	>\$100	>\$200	>\$400	1%	5%
January	log-Normal	93.4738	53.6906	9.2295	0.3015	325.86	233.61
February	log-Normal	93.7169	54.1349	9.2789	0.2967	325.41	233.74
March	log-Normal	94.1881	53.4882	8.1408	0.1995	308.97	224.70
April	log-Normal	93.2310	50.7751	7.2924	0.1705	301.40	218.75
May	log-Normal	92.2258	49.5421	7.4457	0.2018	306.66	220.49
June	log-Normal	90.8520	46.5041	6.5882	0.1713	298.37	213.91
July	log-Normal	89.0990	44.5710	6.6185	0.2032	303.23	214.70
August	log-Normal	88.1348	44.6064	7.3116	0.2800	316.67	221.25
September	log-Normal	88.2993	44.1897	6.9117	0.2411	309.85	217.58
October	log-Normal	88.9500	42.9971	5.7425	0.1455	289.85	206.86
November	log-Normal	90.0102	45.1091	6.3263	0.1679	296.54	211.87
December	log-Normal	91.4582	48.9813	7.7717	0.2434	313.67	223.57

*Probability distribution that show the best goodness of fit index in Table 2

Regarding quantiles, it is expected that with a 5% probability there will be an average price higher than \$233.61 in January. On the other hand, it is expected that with a 1% probability there will be an average price higher than \$289.85 in October (Table 3). To achieve better returns, the sales of coffee bags should be concentrated in the months with a higher probability of higher average prices. The months of January, February, and March are the most recommended for commercialization. According to the data, provided by COOXUPÉ, the reality contradicts the indications, as the highest averages in terms of the number of bags sold occur from July to October.

Storing coffee beans and selling during the off-season is one of the possible strategies for coffee farmers and cooperatives to achieve higher profits and revenues. Administratively, product storage becomes viable when the difference between the expected price and the current price is greater than the sum of the interest rate and the storage cost. It becomes

attractive when it can compensate for its associated risks (Arêdes *et al.*, 2008).

Agreement to Arêdes *et al.* (2008), it is important to assess the profitability and risk of the operation. In their work, they used indicators such as Payback Period (PP), Net Present Value (NPV), Internal Rate of Return (IRR), and the Average Cost Indicator (CMe), calculated based on the cash flow of production, to make decisions regarding the investments in coffee storage adoption. The authors confirm the economic viability of storage and pointed out that the months of January to March, which have a higher probability of higher prices per sack, are better for commercialization. Conversely, the least favorable months for commercialization are from June to November, as they have the lowest probabilities for values exceeding 100 dollars.

4. Conclusions

The Gamma and log-Normal distributions fitted the monthly data of average coffee bags prices. The Log-Normal distribution is more suitable for the probabilistic study of the variable in all months. January, February, and March are the months with the highest probability of higher average values and are therefore the most recommended for coffee trading. Probabilistic modeling is effective and can be practically applied to understand price dynamics and formulate strategies for market participants. Its study can be scaled for other places according to the interest of the study.

The current study has some limitations. Conducting a study with covariates is feasible and enriching, such as distinguishing between the varieties of traded coffee, the harvest season, and other directly correlated economic variables. For future studies, it is possible to perform probabilistic modeling with the aforementioned distinctions, as well as the application of other probability distributions, or even modeling the quantities of coffee bags marketed by the cooperative and comparing the two (or more) variables. It is also possible to maintain the chronological nature of the series and use other modeling approaches, such as time series modeling.

Acknowledgments

Acknowledgments to the Federal University of São Carlos - UFSCar, Center for Agricultural Sciences, the National Council for Scientific and Technological Development - CNPq for granting the Undergraduate Research Scholarship, the COOXUPÉ for the availability of data and information provided, and also to the researcher and rural producer Anselmo Augusto de Paiva Custódio for the contributions with information from COOXUPÉ data.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Albieri, A. C. de S., & Terra, L. A. A. Estratégias de diferenciação da commodity de café. *Future Studies Research Journal: Trends and Strategies*, **14** (1), 1–27 (2022). <https://doi.org/10.24023/futurejournal/2175-5825/2022.v14i1.484>
2. Arêdes, A. F. de, Pereira, M. W. G., Santos, V. F. dos, & Santos, M. L. dos. Rentabilidade e Risco na Estocagem do Café pelos Produtores na Região de Viçosa-MG. *Revista de Economia e Agronegócio*, **6** (2), 235 (2008).
3. Arêdes, A. F. De, Pereira, M. W. G., & Teixeira, E. C. *Previsão De Preço Do Quilo Do Café Arábica: Uma Aplicação Dos Modelos Arima E Garch*. (2009).

4. Bautista, E. A. L., Zocchi, S. S., & Angelocci, L. R. Fitting the generalized extreme value distribution (GEV) to the maximum wind speed data in Piracicaba, São Paulo, Brazil. *Revista Matemática e Estatística*, **22** (1), 95–111 (2004)..
5. Camargo, Â. P. de, & Camargo, M. B. P. de. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. *Agrometeorologia*, **60** (1), 65–68 (2001). <https://doi.org/https://doi.org/10.1590/S0006-87052001000100008>
6. Capitani, D. H. D., & Mattos, F. Measurement of commodity price risk: An overview of Brazilian agricultural markets. *Revista de Economia e Sociologia Rural*, **55** (3), 515–532 (2017). <https://doi.org/10.1590/1234-56781806-94790550306>
7. Carrasco-Gutierrez, C. E., & Almeida, F. M. de M. Modelagem e Previsão do Preço do Café Brasileiro. *Revista de Economia*, 7–27 (2013). <https://doi.org/http://dx.doi.org/10.5380/re.v39i2.23476>
8. Casella, G., & Berger, R. R. *Statistical Inference* (2^o ed). Thomson Learning. (2001).
9. CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. *Análise Mensal Café - Outubro de 2018*.
10. CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. *Análise Mensal Café - Outubro/Novembro de 2019*.
11. CONAB - Companhia Nacional de Abastecimento. *Acompanhamento da Safra Brasileira - Safra 2022*. file:///C:/Users/Admin/Downloads/E-book_BoletimZdeZSafraZcafZ-ZjaneiroZ22.pdf
12. Cooxupé. *Cooxupé Exportação*. <https://www.cooxupe.com.br/exportacao/exportacao/> (2022).
13. Cuaresma, J. C., Hlouskova, J., & Obersteiner, M. Fundamentals, speculation or macroeconomic conditions? Modelling and forecasting Arabica coffee prices. *European Review of Agricultural Economics*, **45** (4), 583–615 (2018). <https://doi.org/10.1093/erae/jby010>
14. Devore, J. L. *Probabilidade e estatística para engenharia e ciências* (9^o ed). Cengage. (2018).
15. Ghoshray, A., & Mohan, S. Coffee price dynamics: An analysis of the retail-international price margin. *European Review of Agricultural Economics*, **48** (4), 983–1006 (2021). <https://doi.org/10.1093/erae/jbab027>
16. Hartmann, M., Moala, F. A., & Mendonça, M. A. Estudo das precipitações máximas anuais em Presidente Prudente. *Revista Brasileira de Meteorologia*, **26** (4), 561–568 (2011). <https://doi.org/10.1590/S0102-77862011000400006>
17. ICO - International Coffee Organization. *Volatilidade Dos Preços Do Café: O Covid-19 E Os Fatores Fundamentais Do Mercado* (2020). <http://www.ico.org/documents/cy2019-20/coffee-break-series-2p.pdf>
18. ICO - International Coffee Organization. *Visão Geral do Relatório sobre o Desenvolvimento do Café da OIC de 2020*.
19. ICO - International Coffee Organization. *Relatório sobre o mercado cafeeiro fevereiro 2022*. 1–10. <https://www.ico.org/documents/cy2021-22/cmr-0222-p.pdf>
20. Liska, G. R., Cirillo, M. Â., Menezes, F. S. de, & Filho, J. S. de S. B. A simplex dispersion model for improving precision in the odds ratio confidence interval in mixture experiments. *Acta Scientiarum - Technology*, **42**, 1–10 (2020). <https://doi.org/10.4025/ACTASCITECHNOL.V42I1.44068>
21. Lucca Filho, E. L. F., Liska, G. R., Santos, J. A., & Matiussi, A. C. Modelagem probabilística de preços máximos da commodity boi gordo para o estado de São Paulo. *Nativa*, **10** (1), 22–31 (2022). <https://doi.org/10.31413/nativa.v10i1.13291>

22. MAPA - Ministério da Agricultura Pecuária e Abastecimento. *Conheça a história do café no mundo e como o Brasil se tornou o maior produtor e exportador da bebida*. (2022). <https://www.gov.br/agricultura/pt-br/assuntos/noticias/conheca-a-historia-do-caffe-no-mundo-e-como-o-brasil-se-tornou-o-maior-produtor-e-exportador-da-bebida>
23. Mehta, A., & Chavas, J. P. Responding to the coffee crisis: What can we learn from price dynamics? *Journal of Development Economics*, **85** (1–2), 282–311 (2008). <https://doi.org/10.1016/J.JDEVECO.2006.07.006>
24. Menezes, F. S., Liska, G. R., Cirillo, M. A., & Vivanco, M. J. F. Data classification with binary response through the Boosting algorithm and logistic regression. *Expert Systems with Applications*, **69**, 62–73 (2017). <https://doi.org/10.1016/j.eswa.2016.08.014>
25. Paiva, C. M. N., de Alcântara, J. N., Campos, R. S., & dos Santos, A. C. Competitividade Do Café No Mercado Internacional: Uma Análise Econométrica. *Agroalimentaria - Universidad de los Andes*, **24**, 43–58 (2018).
26. Pohlert, T. *trend: Non-Parametric Trend Tests and Change-Point Detection*. R package version 1.1.5. (2023).
27. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. (2021).
28. Sá Barreto, R. C., & Zugaib, A. C. C. Dinâmica do mercado internacional de café e determinantes na formação de preços. *Economia & Região*, **4** (2), 7 (2016). <https://doi.org/10.5433/2317-627x.2016v4n2p7>
29. Sá, E. A. S., Moura, C. N. de, Padilha, V. L., & Campos, C. G. C. Trends in daily precipitation in highlands region of Santa Catarina, southern Brazil. *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, **13** (1), (2018). <https://doi.org/10.4136/ambi-agua.2149>
30. Silveira, R. L. F. da, Mattos, F. L., & Saes, M. S. M. The Reaction of Coffee Futures Price Volatility to Crop Reports. *Emerging Markets Finance and Trade*, **53** (10), 2361–2376 (2016). <https://doi.org/10.1080/1540496X.2016.1205976>
31. Teixeira, O. M., Liska, G. R., & SANTOS, J. A. Dos. Modelagem probabilística e temporal do mercado de café no Brasil. *Anais do 61º Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural (SOBER)*. (2023). <https://doi.org/10.29327/sober2023.623960>
32. Venables, W. N., & Ripley, B. D. *Modern Applied Statistics with S* (4º ed). Springer. (2002).
33. Zambrano-Bigiarini, M. *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series*. (2020). <https://doi.org/10.5281/zenodo.839854>